Modelling information quality and source reliability to improve the trust of volunteered geographic information

Paul Goodhue[1]*, Ioannis Delikostidis[1]

[1]Department of Geography, College of Science, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand.

*Corresponding author, E-mail address: paul.goodhue@pg.canterbury.ac.nz (Paul Goodhue).

Key Words: Crowdsourcing, VGI, Quality, Trust, Reliability.

Abstract:

Crowdsourcing and volunteered geographic information (VGI) can improve the way we collect information about the world we live in, but they are not without limitations. Traditionally sourced geographic information (GI) is implicitly trusted due to the authority of its source, whereas VGI often lacks trust. The lack of trust in VGI stems from the very nature of crowdsourcing where many sources of information form a heterogeneous crowd that in turn produces information of heterogeneous quality. Improving the trust of VGI requires knowledge of its quality and the reliability of its sources and collection methods. To improve the trust of VGI, a model is proposed to assess the quality of its spatio-temporal and semantic components and the reliability of the individual producing the information and their collection methods. The proposed model is analysed through a case study with the collection of VGI representing walking and biking tracks. The case study employs the proposed model to generate trust ratings of the collected VGI. Comparisons of the modelled trust and established baseline trust outline improvements in the trust of the VGI through the model. Measurable improvements to the trust of VGI are expected with the introduction of the crowdsourcing model, with varying degrees of improvement across the model components.

1. Introduction:

The concept of crowdsourcing has come to fruition in recent years thanks to the proliferation of web 2.0 and supporting technologies. Crowdsourcing has changed the landscape of information collection and dissemination from a top-down flow of information from producer to users, to a bi-directional model where information users may also be producers (Coleman, Georgiadou, & Labonte 2009; Heipke 2010). A geographic application of crowdsourcing is the concept of volunteered geographic information (VGI) (Goodchild 2007), where geographic information (GI) is actively created by a crowd, in many cases the general public, for a common cause. VGI differs from traditional GI in its potential for many sources of information. Traditional GI is often collected by agencies for their own needs or to on-sell as a product, whereas VGI is collected by individuals within the crowd working towards a common goal (Flanagin & Metzger 2008). Traditional GI sources may be implicitly trusted as they hold authority over the data they are collecting, whereas VGI may lack trust due to its very nature. This paper presents a model to improve the trust of VGI through assessments of its quality and the reliability of its sources and collection methods.

2. Trust in Volunteered Geographic Information:

Trust of VGI is in knowing the quality of the information and the reliability of its source. VGI quality is a combination of its spatio-temporal and thematic accuracy, completeness and consistency (Haklay 2010). Trust may be based upon understanding its spatial, temporal, and thematic accuracy through quality assurances and assessments (Bordogna et al. 2013; Goodchild & Li 2012; Ostermann & Spinsanti 2011). With all GI, assessing the quality relies on predetermined quality benchmarks specific to the domain and purpose of the information. Defining quality benchmarks for VGI is challenging as the purpose of the VGI may not be known and the domain may be broad, leading to a lack of knowledge of the quality of the VGI and in turn a lack of trust. Many VGI applications assure the information quality through concepts such as Linus' Law (Goodchild & Li 2012) and consensus crowdsourcing (Kamar & Horvitz 2012), both of which require a large and active crowd. VGI quality may also be assured by constraining the information that can be collected, but knowledge that does not fit within the constraints may be lost. In addition to quality, trust in VGI requires reliable information sources and the collection methods. VGI sources are volunteers with varying levels of expertise and motivation, forming crowds that are heterogeneous in nature, that in turn produce VGI that is heterogeneous in quality (Bordogna et al. 2014; Elwood 2008; Haklay et al. 2010). As VGI sources are often numerous and unknown, the reliability of the sources must be explicitly measured to improve VGI trust. The reliability of a VGI source can be determined through their reputation within the VGI project, or expertise in the VGI domain, both of which may be difficult to quantify. Many VGI projects make use of source reputations to improve VGI trust, such as the USGS project "The National Mapping Corps", where individuals are assigned levels of authority based on their quantity of data submissions (McCartney, Bearden, & Newell 2013). Understanding the impacts of collection methods on VGI quality also improves VGI trust. For example, the collection of VGI through a smartphone application can yield varying spatial accuracies depending on the location capture technique used, where the differences between the accuracy of an Assisted Global Positioning System A-GPS feature and one derived from a cellular network may be in the order of hundreds of metres (Zandbergen 2009). As improving the trust of VGI requires knowledge of its quality and source reliability, robust quality and reliability assessments are required. Assessments of

the spatial, temporal, and thematic components of VGI may be derived from assessments of traditional GI, whereas assessments of the reliability of the source and collection methods are specific to VGI. Many VGI examples in the literature contain project specific methods to improve VGI trust that are driven by the project domain and purpose.

3. Proposed Model:

The proposed model aims to improve the trust of VGI through assessments of the quality of VGI and the reliability of its source and collection methods. In doing so, the model is intended for use with unconstrained VGI to ensure knowledge is not lost through the VGI collection. The quality of VGI is determined through assessments of its spatio-temporal and semantic components describing where, when and what the feature represents. The reliability of the source and capture methods are determined through assessments of metadata collected alongside the VGI detailing who collected the VGI and how it was collected. This gives the model four main components; spatio-temporal, semantic, social, and methods. The proposed model does not aim to improve the quality of VGI, but through knowing the VGI quality and having trust in the VGI, potential users of the information will have confidence in its limitations.

4. Case Study:

The ability of the proposed model to improve the trust of VGI is assessed through a case study. The context of the case study is VGI representing walking and biking tracks; chosen as it has the potential to create spatio-temporally and thematically rich VGI about entities the general public actively use. To ensure rich VGI is collected for the case study, a corresponding, detailed data model has been established. The base data model employs the Local Government Geospatial Alliance's (LGGA) National Tracks Schema (Local Government Geospatial Alliance, 2014), designed to normalise walking and biking track information across New Zealand. Expanding upon the National Tracks Schema, the data model includes fields that represent additional information about the VGI feature, for example the activity the individual is using the track for, or metadata about the VGI source and collection methods. Complex geometries including multiline features and points of interest recorded alongside the track feature are also catered for in the data model. The case study collects VGI through Android and web applications created specifically for the project. These platforms were chosen to allow the crowd to collect VGI in the field through Global Positioning System (GPS) and also edit the VGI and digitise new features. The multiple VGI collection methods provided by the applications allow for the collection of metadata on the use of each application and their influences on the trust of the VGI. Additionally, the applications collect metadata on the source of the VGI such as the individual that originally created a feature and the individual(s) who edited the feature. This information will construct reputations of the VGI sources to be assessed by the social component of the crowdsourcing model. Along with the crowdsourcing application, the case study employs a crowdsourcing model to assess the four main trust components of VGI; spatio-temporal, semantic, social and methods. The crowdsourcing model assesses the VGI and metadata by comparing and contrasting it to related information, generating quantified trust ratings for each assessment and trust component of the VGI. The analysis portion of the case study determines the extent to which the crowdsourcing model components improve the trust of the VGI. These assessments utilise

benchmarks of quality for each model component determined through quality elements of authoritative track GI, and a baseline trust value of the VGI. To determine the baseline trust of the VGI, the result of each assessment is observed individually as a baseline trust rating and compared to the modelled trust rating comprised of the ratings from the remaining assessments. This approach is employed to outline any improvements in the trust of the VGI through the use of the crowdsourcing model and to ensure self-referencing trust ratings are not introduced into the analysis.

5. Expected Results:

Measurable improvements to the trust of VGI are expected with the introduction of the crowdsourcing model, with varying degrees of improvement across the model components. It is hoped that improvements to VGI trust supplied by the model components coupled with the cost and complexity of each component will form an understanding of each component's value.

6. Conclusion:

Crowdsourcing has the potential to improve the way we collect information about the world we live in, but it is not without limitations. Due to the very nature of crowdsourcing and VGI there are often many sources of information from a heterogeneous crowd, resulting in a lack of trust of the information. Improving the trust of VGI requires knowledge of the VGI quality and the reliability of its sources and collection methods. This paper presented a model to improve the trust of VGI through assessments of the four main trust components of VGI; spatio-temporal, semantic, social, and methods. The presented case study aims to collect VGI through Android and web applications to analyse the extent to which the proposed model improves the trust of VGI, outlining the value of the model components.

7. Acknowledgements:

We gratefully acknowledge the support of the Cooperative Research Centre for Spatial Information (CRCSI).

8. **References:**

Bordogna, G., Carrara, P., Criscuolo, L., Pepe, M., and Rampini, A. (2014). A linguistic decision making approach to assess the quality of volunteer geographic information for citizen science. Information Sciences, 258, 312-327.

Coleman, D. J., Georgiadou, Y., and Labonte, J. (2009). Volunteered geographic information: The nature and motivation of produsers. International Journal of Spatial Data Infrastructures Research, 4(1), 332-358.

Elwood, S. (2008). Volunteered geographic information: future research directions motivated by critical, participatory, and feminist GIS. GeoJournal, 72(3-4), 173-183.

Flanagin, A. J., and Metzger, M. J. (2008). The credibility of volunteered geographic information. GeoJournal, 72(3-4), 137-148.

Goodchild, M.F. (2007). Citizens as sensors: the world of volunteered geography. GeoJournal, 69(4), 211-221.

Goodchild, M. F., and Li, L. (2012). Assuring the quality of volunteered geographic information. Spatial Statistics, 1, 110-120.

Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. Environment and Planning B: Planning and Design, 37, 682-703.

Haklay, M., Basiouka, S., Antoniou, V., and Ather, A. (2010). How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information. The Cartographic Journal, 47(4), 315-322.

Heipke, C. (2010). Crowdsourcing geospatial data. ISPRS Journal of Photogrammetry and Remote Sensing, 65, 550-557.

Kamar, E., and Horvitz, E. (2012). Incentives for truthful reporting in crowdsourcing. In Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems, 3, 1329-1330.

Local Government Geospatial Alliance. (2014). LGGA Tracks Project. Retrieved February 15, 2017 from http://lgga-nz.blogspot.co.nz/p/tracks-project.html.

McCartney, E., Bearden, M., and Newell, M. (2013). Crowd-Sourcing the Nation: Now a National Effort. Retrieved February 15, 2017 from https://archive.usgs.gov/archive/sites/www.usgs.gov/newsroom/article.asp-ID=3664.html.

Ostermann, F. O., and Spinsanti, L. (2011). A Conceptual Workflow For Automatically Assessing The Quality Of Volunteered Geographic Information For Crisis Management. Proceedings of the 14th AGILE International Conference on Geographic Information Science.

Zandbergen, P. A. (2009). Accuracy of iPhone locations: A comparison of assisted GPS, WiFi and cellular positioning. Transactions in GIS, 13(s1), 5-25.